being re-projected, wherein the temporary objects are in front of the face being re-projected for less than a threshold period of time.

Some implementations can include a non-transitory computer readable medium having stored thereon software instructions that, when executed by one or more processors, cause the one or more processors to perform operations. The operations can include receiving a video including a plurality of frames, the video captured by a physical camera at a first point of view, wherein the video includes depth information corresponding to the plurality of frames, and detecting a face within the video, wherein the face is within a foreground portion of one or more frames of the video. The operations can also include determining the foreground portion of the plurality of frames based on one or more depth values of the depth information corresponding to the plurality of frames, and positioning a virtual camera at a second point of view, wherein the second point of view is different from the first point of view.

The operations can further include obtaining a projection matrix of the foreground portion based on the virtual camera, the projection matrix corresponding to the second point of view, and generating a modified video that includes a modified foreground portion based on the projection matrix. The operations can also include stabilizing the modified foreground portion of the modified video. In some implementations, determining the foreground portion can include detecting the face and a background, where the detecting includes extruding the foreground portion of the face from the background.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is a diagram showing an example user device being used to conduct a video call with the user device capturing the user at an angle that may produce a lower quality video;

FIG. **2** is a diagram showing an example user device being used to conduct a video call with the user device capturing the user at an angle that may produce a lower quality video and a virtual point of view used to correct the pose of the user to produce a higher quality video in accordance with some implementations;

FIG. **3** is a diagram of an example user device showing a video call in which the video was captured at an angle that may produce a lower quality video;

FIG. **4** a diagram of an example user device showing a video call in which the video was re-projected to create a corrected pose and a higher quality video;

FIG. **5** is a flowchart of an example method for pose correction in accordance with some implementations;

FIG. **6** is a flowchart of an example method for pose correction in accordance with some implementations; and

FIG. **7** is a diagram of an example computing device configured for pose correction in accordance with some implementations.

DETAILED DESCRIPTION

Some implementations include methods and systems to modify or correct a user pose in a video call or conference. Users may be participating in a video call with a camera at an angle that may produce a lower quality video or moving to such an angle at some point during a call. For example, a user may appear to be looking up in a video based on how the user is holding their phone. Most video calling devices, e.g., smartphones, laptops, tablets, etc. currently include a camera at the top end of the device. Such placement often produces images that are less than optimal for a video call. Images suitable for pose correction by an implementation can include single images such as photos, cinemagraphs and other animated images, videos, etc.

Some implementations can create an adjusted video call presence by re-projecting the video image of the user using a virtual point of view (or virtual camera angle) such that the user appears to be at an angle that produces a corrected pose (e.g., directly facing the camera, etc.). Some implementations can also include correction for camera distortion (e.g., to remove the big nose effect), reduction in the effect of jitter (e.g., caused by user hand movement), etc.

By capturing color and depth information from video frames, an implementation can extrude depth for foreground. The system can place a virtual camera at a different angle than the physical camera of the device (e.g., up to give the effect that others are looking at the user directly).

Some implementations can include three main operations: 1) in-frame face detection; 2) pose correction; and 3) gaze correction. In some implementations, in-frame face detection can include using technology that provides a face pose angle (e.g., a 3D Euler angle) and the position of face in a frame (such as ARKit or ARCore, which are trademarks of their respective owners). The pose angle can include the angle of the face relative to the camera (e.g., looking straight ahead is 0,0,0, and when face is rotated about one of the axes, then the angle corresponding to that axis changes in the 3D Euler angle). The position is the x, y position within the color image, from which the device can determine position relative to the device (e.g., x, y, z position). Extra padding can be added to the frame to provide a buffer. Whenever the face of the user is within a threshold (or safe) range of angle relative to the camera, (e.g., plus or minus about 20 degrees around the x or y axes) the system returns a high confidence value indicating that the face of the user is positioned within a frame of video such that the techniques for pose correction described herein are most likely to be effective. If the image of the user's face is in the padding area, confidence is lower (e.g., indicating that the pose correction techniques described herein may not work as effectively given the user's face position), and decreases as the portion of the face in the padding area increases. No confidence (or a zero or near zero confidence value) is returned if the user is outside of the acceptable range (e.g., the frame and/or padding).

Confidence, reason or basis, and/or guidance to a user can be provided based on the in-frame face detection. As discussed above, confidence indicates a level of confidence that the pose correction technique can be applied effectively to a given face position within a frame of video or other image. Reason or basis can include an indication of why the confidence score may be low (e.g., face too close to a border, etc.). Guidance can include an indication provided by the system (e.g., via a graphical user interface) to help direct the user to correct the issue that has caused a low confidence score. For example, if confidence is low, a reason can be determined and feedback based on the reason can be provided to the user (e.g., a black overshadow that increases in transparency with lower confidence, or a small arrow directing the user to move back in frame, etc.) to help guide the user to change the video capture to increase confidence.

When the image capture device provides depth data, e.g., real time depth data captured using a depth sensor or other hardware of the device, implementations can take into account such data along with the face pose. A camera RGB feed can be represented as a plane that is perpendicular to the camera forward (e.g., a vector pointing out of the lens of the camera in a forward direction toward a subject the camera is